EXHIBIT 10

BROWN DECLARATION IN SUPPORT OF MOTION TO EXCLUDE

inference about a coefficient if and only if the estimated coefficients and standard errors are the same for all specifications.

The last problem is similar:

Problem 5. What measure of overall confidence analogous to an R^2 should apply to a research effort which reports many different equations with different R^2s? We subsequently propose a special kind of average R^2. Again, there is no classical counterpart.

In the first two sections of this chapter, the problem of identifying the class of admissible tests is distinguished from the problem of selecting a particular test. Classical hypothesis testing concerns itself almost exclusively with the first problem, but it has nothing meaningful to say about the second. The rule of thumb quite popular now, that is, setting the significance level arbitrarily to .05, is shown to be deficient in the sense that from every reasonable viewpoint the significance level should be a decreasing function of sample size.

A few words may now be said in anticipation of the sections to follow, which describe in detail the Bayesian approach to hypothesis testing. By Bayes' rule, the relative posterior probabilities of two hypotheses can be written as

$$\frac{P(H_i|\mathbf{Y})}{P(H_j|\mathbf{Y})}=\left[\frac{P(\mathbf{Y}|H_i)}{P(\mathbf{Y}|H_j)}\right]\left[\frac{P(H_i)}{P(H_j)}\right]. \tag{4.4}$$

The second factor in brackets is the prior odds ratio in favor of H_i. The data-dependent term in the first set of brackets is the "Bayes factor."

The data are said to favor H_i relative to H_j if the Bayes factor exceeds one, that is, if the observed data $\mathbf{Y}$ is more likely under hypothesis H_i than it is under hypothesis H_j. The densities of $\mathbf{Y}$ implied by the hypotheses (4.1) are conditional on the parameters, $\boldsymbol{\beta}_i$ and σ_i^2, but may be straightforwardly "mixed" into a marginal density as

$$f(\mathbf{Y}|H_i)=\int_{\boldsymbol{\beta}_i}\int_{\sigma_i^2} f(\mathbf{Y}|H_i,\boldsymbol{\beta}_i,\sigma_i^2)\,f(\boldsymbol{\beta}_i,\sigma_i^2)\,d\sigma_i^2\,d\boldsymbol{\beta}_i \tag{4.5}$$

where $f(\boldsymbol{\beta}_i,\sigma_i^2)$ is the prior density. The conditional p.d.f. $f(\mathbf{Y}|H_i,\boldsymbol{\beta}_i,\sigma_i^2)$ for a particular value of $\mathbf{Y}$ is a likelihood function of $(\boldsymbol{\beta}_i,\sigma_i^2)$, and (4.5) defines $f(\mathbf{Y}|H_i)$ as a weighted or marginal likelihood.

The Bayes factor is to be contrasted with the likelihood ratio, which is used classically to summarize the data evidence. The likelihood ratio is

$$L(H_i,H_j)=\frac{\max\limits_{\boldsymbol{\beta}_i,\sigma_i^2} f(\mathbf{Y}|\boldsymbol{\beta}_i,\sigma_i^2,H_i)}{\max\limits_{\boldsymbol{\beta}_j,\sigma_j^2} f(\mathbf{Y}|\boldsymbol{\beta}_j,\sigma_j^2,H_j)}.$$

The Bayes factor averages the likelihood function over all values of $(\boldsymbol{\beta}_i,\sigma_i^2)$. The likelihood ratio evaluates the likelihood function at its maximum.

Any attempt to summarize the data evidence in favor of the hypotheses (4.1) leads to an irreconcilable index number problem of the following form. If $\boldsymbol{\beta}_i$ assumed one value, the data evidence could be said unambiguously to favor the ith hypothesis, but if $\boldsymbol{\beta}_i$ assumed another value, the data unambiguously cast doubt on H_i. Since H_i allows $\boldsymbol{\beta}_i$ to assume any value, the data evidence is necessarily ambiguous.

The classical solution to this dilemma seems most appropriate for testing a point null hypothesis against a composite alternative. The null hypothesis is regarded as the favorite; it is the one that is being "tested." If there is *any* way for the alternative to look as good as the null hypothesis, we should be worried about retaining the null as the favorite. Consequently, we identify the evidence *against* the null hypothesis in terms of the evidence in favor of the alternative at the value of $\boldsymbol{\beta}$ that makes the alternative appear best. The appropriate statement however is not that the alternative is favored. All that is said is that the alternative is *conceivably* favored.

There is a great tendency in practice to forget the all-important word "conceivably" in this sentence, and as a consequence, classical tests distort the data evidence. In the more common case when the null hypothesis is a composite hypothesis, classical tests usually also evaluate the data evidence at the parameter point that makes the null hypothesis appear best. The resulting statement about the evidence is: "If each hypothesis is allowed to 'put its best foot forward,' hypothesis H_j is favored." In practice, the qualifying phrase "if...forward," is often forgotten, and the data evidence may consequently be significantly distorted.

A Bayesian approach, in contrast, presupposes a prior distribution that can be used to weight the evidence at different values of the parameters. Thus instead of letting an hypothesis "put its best foot forward," the performance at all values of the parameters is considered. The apparent problem that then arises is the construction of a nonarbitrary weight function. Here and elsewhere, we take the position that a researcher is obligated to report as fully as possible the mapping of priors into posteriors. He should describe the data evidence as favoring hypothesis H_1 if the prior takes one form, and favoring H_2 if it takes another. He thereby avoids having to make a choice that rightly belongs to his readers: the choice of prior distribution.

4.1. Hypothesis Testing: A Judicial Analogy

The subject of hypothesis testing may be usefully introduced by an analogy. Based on the evidence presented, a judge and jury in a legal

proceeding decide whether a defendant should be set free or sent to jail. If they decide that the evidence favors the hypothesis of guilt, they accordingly send the defendant to jail. Otherwise he is set free. The assumption of innocence until proven guilty beyond a reasonable doubt explicitly favors the hypothesis of innocence. We refer to this favored hypothesis as the *null* hypothesis or H_0 and the hypothesis of guilt as the *alternative* hypothesis or H_1. The evaluation of the evidence and the decision either to free or jail the defendant is called a "test" of the null hypothesis *against* the alternative, and the decision is described as acceptance versus rejection of the hypothesis of innocence.

The more critical error—sending an innocent man to jail—is called an *error of the first kind* or *type I error*. Acceptance of the null hypothesis when it is in fact false—freeing a guilty man—is called an *error of the second kind* or *type II error*. Schematically we have

Actions

Hypotheses (States)		Set Free (accept H_0)	Send to Jail (reject H_0)
H_0:	Innocent		Type I error
H_1:	Guilty	Type II error	

If a man is innocent, we want to have a low probability of sending him to jail. Let this probability be α

$$\alpha = P(\text{jail}|\text{innocent defendant}).$$

Analogously, let

$$\beta = P(\text{set free}|\text{guilty defendant}).$$

Both α and β are defined before the judicial process commences. In effect, they predict the quality of the evidence and the ability of the court to process the evidence effectively. For example, a low value of α amounts to the prediction that if the defendant is innocent, the evidence will be so unambiguous and the process by which a verdict is rendered will be so perfect that with near certainty he will be justly found innocent.

The theory of hypotheses testing deals with defining procedures such that α and β are small. A typical choice set for α and β is depicted in Figure 4.1. Flipping a coin to decide whether to free or jail the suspect implies $\alpha = \beta = .5$. The line running from the point $(\alpha = 1, \beta = 0)$ to $(\alpha = 0, \beta = 1)$ represents the set of all such randomized decisions. The value $(\alpha = 0, \beta = 0)$ represents perfect evidence and a perfect procedure which is excluded

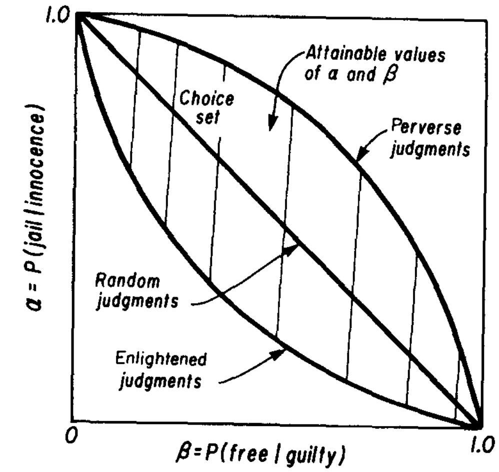

Fig. 4.1 Probabilities of error.

from the choice set in Figure 4.1. If $(\alpha = 0, \beta = 0)$ is not available, the perfect error point $(\alpha = 1, \beta = 1)$ is similarly not available, since if we could be sure of making an error, by doing the exact opposite we could be sure of *not* making an error. The curved line labeled "enlightened judgments" represents the best possible court procedures based on the available evidence. The curve labeled "perverse judgments" is just the mirror image of the "enlightened judgment" curve, involving the exact opposite action.

The choice of a courtroom procedure is usefully thought to involve two steps. The first step is to identify the set of procedures that involve enlightened use of the evidence, that is, those that make α and β as small as possible. The second step is to choose a particular procedure from among this set of admissible procedures. The former is a logical mathematical problem that admits a clear-cut uncontroversial solution; the latter is not. Let us consider the latter problem.

The essential problem the court faces once the line of enlightened judgments is computed is that stricter interpretation of the given body of evidence and a greater tendency to send men to jail which could reduce the probability β of freeing guilty men necessarily increases the probability α of jailing innocent men. By assumption, it is desirable to have both α and β small. The choice dilemma is that reduction of one necessarily leads to an increase in the other. Actual choice can be said to depend on a preference function $U(\alpha, \beta)$ indicating numerically the level of satisfaction attained if the courtroom procedure yields probabilities α and β. Several "contour" lines of a typical preference function are indicated in Figure 4.2.